Han, B. M., Zhu, X.N., Wong, W.G., **Ferreira, L**. and Teng, J. (2002). Study into key problems in forecasting export containers in a region of China. In*:* Proc of International conference on Traffic and Transportation Studies, Wang, K. C. P., Xiao, G. Nie, L. and Yang, H. (Eds), Vol 1, 274-285. American Society of Civil Engineers, Virginia, USA.

Study into key problems in forecasting export containers in a region of China

Abstract: Forecasting network data traffic is an important part of the function of planning and managing information systems. However, the contents of network data are so stochastic and complex that it is very difficult to establish stable functions to describe the mapping relationship between data flows and associated causal influences. In this paper, a multi-layer feed forward neural networks (NN) model is put forward to identify such relationship and the corresponding learning rule of NN, back-propagation (BP) algorithm, is given. In addition necessary estimation and validation processes are designed to ensure the successful implementation of the model proposed. The paper elucidates the application of NN model around the case of forecasting China west railway Transportation Management Information Systems (TMIS) network traffic. The predictive results obtained demonstrate that the NN model and the solution algorithm are very applicable for information planning on the TMIS network in west China.

Keywords: Network data traffic, TMIS, neural networks, and back-propagation

1. Introduction

In the telecommunications industry, planning network-loading capacity has become a focus, which usually includes two aspects of studies. One aspect aims at optimizing the costs of capacitated facilities of a network. For example, Magnanti (1995) develops modeling and solution approaches for loading facilities to satisfy the given demand at minimum cost. Anantaram (1995) develops a decomposition algorithm for local access telecommunications network expansion planning based on a minimum cost optimization methodology. Their studies are done based on known planning demand of network data traffic. The second aspect of past studies has been forecasting network data traffic, which is an important and practical work for planning future communication networks. This aspect can be divided into two sub-areas of research, namely: forecasting market service demand for a network. For example, Wright (1998) presents a case study of the application of the Delphi Method for forecasting the market for broadband telecommunications in the year 2000. He evaluates market demand from seven different viewpoints and the results were obtained for the market in the whole of North America with a focussed case study of the Toronto urban core. The other sub-area is forecasting network data traffic through mining the mapping relationships between network service demand and network data traffic. However, this latter problem is very complex, because many network services work together simultaneously and almost all the services are real-time and stochastic. It is very difficult to find stable functions describing the above mapping relationships. In this paper, we try to apply neural networks to identify such mapping relationships and sequentially forecast the network data traffic based on future network service demand.

In the following sections, we will firstly introduce some background related to TMIS and then raise the specific question to be solved in this paper. In section 2 the characteristics of the problem and the NN model are discussed and analyzed. Section 3 describes the framework of the NN applied in this paper and the BP algorithm is simply reviewed. In order to select the "best " hidden-node layer scheme, a series of steps mainly including estimation and validation processes are described in section 4. In section 5 the model, the algorithm, estimation and validation processes are applied together in a practical case. The results of the case are also analyzed. the main conclusions are offered in section 6 and some issues which will need further research are highlighted.

1.1 An introduction to Transportation Management Information Systems (TMIS)

Transport infrastructure plays a pivotal role in the economic development of any region, particularly in a developing country context. Since the Chinese government implemented the strategy "China West Development" in 1999, more than 130 billion RMB Yuan has been invested to build western infrastructure, including a high priority given to railway investments. As important as the western railway investment programme, has been the development of railway information engineering, the goal of which is to realize Chinese railway modernization and sustainable development.

TMIS is the main component of Chinese railway information Engineering and was implemented progressively since 1992 By 1999, TMIS project had seen an investment of 2.1 billion RMB Yuan by the Chinese government and 47 million $US in foreign capital.

During the period of Chinese "the Ninth Five Years Plan" (from 1994 to 1999), the main seven sub-projects of TMIS had been completed. These are management information systems for: basic data, dispatching data, real-time vehicle-tracking data, freight transportation, passenger transportation, intermodal transportation and integrated transportation. The basis architecture of TMIS is shown in Figure 1.

The main data saved, processed and transmitted in TMIS comes from the entire processes related with railway transportation management and operation. The information flow extracted from the data flow can be used for aiding different management decision-making levels. Currently, in the Chinese Ministry of Railways (MoR), there are four transportation management/operation levels, which are MoR, railway bureau, railway sub-bureau, and railway station and service section. The railway transportation equipment is almost all distributed in the lower level, namely in the stations and service sections. Railway sub-bureaux mainly fulfill the direct management and control of transportation production. Several railway sub-bureaux make up a railway bureau, which is usually in charge of a regional network. The transportation command center of MoR harmonizes railway bureaux' operations and balances all the vehicle cars' distribution. Between each pair of adjacent longitudinal levels, there exist data flows in dual directions in TMIS. In addition, in order to realize trains formation, running and breakup processes, cargo loading and unloading processes, and passenger boarding and alighting processes, transverse data flows exist between the units or departments in the same level. Figure 2 simply shows the longitudinal and transverse data-flow processes.

TMIS is the most important part of the Chinese railway information engineering, not only because it covers transportation production, but also because the basic data managed by TMIS needs to be shared with the other information systems, such as statistics and analysis system, accounting system, electronic commerce system, railway GIS (Geographical Information System) and others. After the China east TMIS was completed, the above series of information systems have been planned and implemented. In Chinese "the Tenth Five year plan" (From 2000 to 2004), the chief goal of railway information engineering has been decided, which requires that in the next phase, the work of the TMIS project is to enhance the functions of each sub-system and to extend TMIS to the overall railway network. The later is also a process from prototype to practice. In addition, other nine MIS projects paralleling TMIS, including fifty-five special topics will be built simultaneously. In particular, the principle is emphasized that the other systems being built must be harmonized with TMIS and must not hinder the smooth operation of TMIS.

1.2 Question specification

The rate of economic development and the growth in the regional economy translates directly in increases in demand for freight rail transport. With the expected rapid development of west China, railway freight traffic will experience a significant increase. From the TMIS introduction we can see that the information service demand coming from transportation production is the key and direct element influencing network data traffic transmitted in TMIS. What is the relationship between the network data traffic and the service demand? How should the future TMIS network data traffic be estimated? Only by solving these questions, can the planner estimate the capacity of the TMIS communication network.

2. Methodology

The main components of the TMIS network data flows are shown in Tables 1 and 2. In the entire three data exchange platforms (Figure 2), the data traffic transmitted in platform I is the biggest, because the sub-bureau level is directly in charge of dispatching trains and directing vehicle running, which needs real-time supervision. In this platform, the situation of passengers' and arrivals and depqartures in each time period needs to be successively reported to the relevant sub-bureau from the stations. Axle temperature detecting information and vehicle and engine code-scanning information must be transmitted for the management of the database in real-time. Therefore, the information exchange in railway transportation production takes place mainly in platform I.

The data exchange relationships in the other two levels are almost the same as in platform I (see table 1 and 2), but the data contents are different. Data in these two levels are mostly the statistical results and need not be transmitted in real-time. The data traffic transmitted in platform © is basically in proportion to the number of railway bureaux. The amount of data transmitted in level ® is basically in the proportion to the number of railway sub-bureaux. However, the data traffic transmitted in platform I is not in proportion to the number of railway stations and railway service sections. In this paper we only solve the problem of data traffic transmitted in platform I. The problem related to the two other platforms can be solved easily once the first question is solved.

In data exchange platform I, every minute there are so complex data transmitted. On the side of sub-bureau, dispatchers will communicate with stations or sections frequently. For example, the train dispatcher needs to issue dispatching orders to station staff and obtain information from the stations related to current vehicle status. If there is a marshalling station involded, the progress of trains being assembled must also be communicated. The freight transportation dispatcher needs to deal with freight cars distribution, knowing the progress of unloading and loading and freight flow status. The work progress of any production phase must be sent to the corresponding dispatchers and shared with relevant stations and sections at regular intervals. From the above, it can be seen that there exists many stochastic data exchanges in the transportation operation. In addition, the activities made by different people when dealing with the same work may be different. Hence, it is very difficult to define a function to describe the relationship between data traffic transmitted in real-time and the associated practical operation.

However, all TMIS functions are designed around the transportation production. Therefore, in a relative long period there may be some relationship between the sum data traffic transmitted and production scale or output of railway transportation. We hope to find the stable mapping relationship in data exchange platform ∇. If the stable mapping relationship is identified, we can forecast future network data traffic in the communication network according to the planning production scale and transportation output. Future communication network design capacity can be estimated based on the increasing rate of network data traffic.

, It was decided to apply a multi-layer feed forward NN to form this forecasting model. NN has the ability to approximate a desired mapping relationship from training samples, in particular when there exist many complex influences and perhaps both linear and nonlinear relationships are present. Repeatedly training endows NN with the ability of "a black box".

3. Model Formulation

3.1 Model framework

The model applied described here is a two-layer fully connected feed forward NN, which has one input layer with multi-nodes representing independent variables; one hidden layer with sufficient multi-nodes; and one output layer with only one output node representing the dependent variable. Figure 3 shows the model architecture. Each node in the input layer receives a corresponding value of an input independent variable (x_i). The input independent variables are the factors influencing the network data traffic transmitted in the communication network. The only dependent variable (y_h^2) output from the output layer represents the network data traffic transmitted.

Firstly, we give the general equation of a fully connected feed forward NN with *K* layers, we can assume that the input layer is 0th layer and other layers from the first hidden layer to the output layer are coded from 1 to *K* in turn. n_l is the node number of layer *l* and $l = (0,1\cdots K)$. Then initial input vector (*X*), the weight matrix (W_1), receiving vector (Z_1) and output vector (Y_1) from 0th layer to 1st layer, can be expressed respectively as:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_0} \end{bmatrix},\ W_1 = (\omega_{ij}^1)_{n_0 \times n_1},\ Z_1 = \begin{bmatrix} z_1^1 \\ z_2^1 \\ \vdots \\ z_{n_1}^1 \end{bmatrix} = W_1^T X,\ Y_1 = \begin{bmatrix} y_1^1 \\ y_2^1 \\ \vdots \\ y_{n_1}^1 \end{bmatrix} = f(Z_1). \tag{1}$$

And the weight matrix (W_k), input vector (Z_k) and output vector (Y_k) from (*k*-1)th layer to *k*th layer are expressed respectively as:

$$W_k = (\omega_{jh}^k)_{n_{k-1} \times n_k},\ Z_k = \begin{bmatrix} z_1^k \\ z_2^k \\ \vdots \\ z_{n_k}^k \end{bmatrix} = W_k^T Y_{k-1},\ Y_k = \begin{bmatrix} y_1^k \\ y_2^k \\ \vdots \\ y_{n_k}^k \end{bmatrix} = f(Z_k), \text{ and } k = (2\cdots K). \tag{2}$$

Here a Sigmond function is selected as the activation function f, which can be expressed as:

$$f(z) = \frac{1}{1+e^{-z}}. \tag{3}$$

3.2 BP algorithm review

Before NN application, the most important work is to endow NN with the identification function, which is the key to estimate the NN model. The procedure of NN learning such function is obtained through adjusting weights with sufficient samples. Therefore, a good learning rule is needed. Usually, the back-propagation (BP) learning rule is a valid method to calibrate the weights of a multi-layer feed forward NN. Here we assume that there are *T* samples namely *T* pairs of known input and output vectors ($\{X(t), D(t)\}$) used to train NN. Besides inputting layer coded with 0, there are *K* layers coded with 1 to *K*. Then BP learning or training steps can be simply shown as in Figure 4.

Where:
t is the code of samples or code of learning steps, $t = 1,2,\cdots,T$,
$X(t) = \{x_1(t), x_2(t),\cdots, x_{n_0}(t)\}$,
$D(t) = \{d_1(t), d_2(t),\cdots, d_{n_K}(t)\}$
$W(1) = \{W_1(1), W_2(1),\cdots, W_K(1)\}$ are the initial weights stochastically selected,
W_{T+1} are the final weights outputted after being trained with *T* samples,
ε_t is a coefficient denoting learning efficiency of step *t* (here, ε_t is set as 1),
Other parameters are determined as in conditions (1), (2), and (3).

In each step (*t*), the process of weights adjusting begins from the last layer ($W_K(t)$) and ends at the first layer ($W_1(t)$). Therefore, the learning of network weights with all the training sets can be seen as a backward recursive procedure, which is also a characteristic of the BP algorithm.

4. Model specification

4.1 Setting input and output valuables

The dependant variable output from the NN is the TMIS network data traffic, which can be directly measured from the ports of the servers working for the data exchange in platform I. The independent valuable input to NN are shown in Table 3. The influencing factors were selected according to the following two principles: the factors should be easily quantified and they should reflect the scale and output of railway transportation production.

In Table 3, the independent variables whose code is from one to eight mainly reflect the routine phase dispatching planning and statistical information. Whatever the station scale,, the data size and format of the report handed up and the plan handed down are the same. The frequency of exchange between sub-bureaux and stations or sections in a given period are also the same for the same task. The independent variables whose code is from nine to fourteen are selected to reflect the accumulation of the real-time data traffic..

4.2 NN implementation

Figure 5 shows the four steps used in the NN implementation. The main work is to decide the node number in the hidden-layer. In general practice, trial and error is usually applied (Mozolin 2000). Namely, we first set several candidate schemes (referred to as hidden-node schemes), each of which has a different number of nodes in the hidden layer. A series of steps will be performed to select the most adaptive one. All the hidden-node schemes must be processed with the same sample set selected randomly from the full data set and compared under the same criteria.

4.2.1 Training process

After setting the hidden-node schemes, the first step is the training process, whose goal is to minimize the total error for all examples in the training set (Figure 5). The learning algorithm, BP algorithm, needs to be used to adjust the weights set (*W*) recursively. For the procedure of weights adjusting, both the input vectors and actual output vectors of samples can be seen as input. After the last sample is applied to train the NN, the calibrating process is completed. Here we assume that there are *P* sample sets selected stochastically from the full data set to train the NN. In each sample set, there are *N* pairs of known input and output vectors. Each hidden-node scheme will be trained respectively and individually with the *P* sets. Following the weights calibration, each scheme has *P* weights sets ($W_1, W_2, \cdots, W_P$).

4.2.2 Estimation process

In the second step, the same *P* sample sets (Input A) used in the training process will be input into the NN again (Figure 5). This time the NN weights set has been calibrated and the error between the practical output and the actual output can be calculated. We will select the best-qualified scheme by comparing the errors of all the schemes.

For each scheme, the hidden-node scheme evaluation function can be expressed as equation 4, 5 and 6. The scheme with the smaller $G(W)$ and $G_{\max}(W)$ value will be selected.

$$G(W_i) = \left[\frac{1}{N}\left(\frac{D_i - Y_i}{D_i}\right)^{\mathrm{T}}\left(\frac{D_i - Y_i}{D_i}\right)\right]^{1/2}, \tag{4}$$

$$G(W) = \frac{1}{P}\sum_{i=1}^{P} G\ (W_i), \tag{5}$$

$$G_{\max}(W) = \max\left(\left|\frac{d_{ij} - y_{ij}}{d_{ij}}\right|\right),\ i \in (1 \cdots P) \text{ and } j \in (1 \cdots N). \tag{6}$$

Here $D_i = (d_{i1}, d_{i2}, \cdots, d_{ij}, \cdots, d_{iN})$ is the actual output vector and $Y_i = (y_{i1}, y_{i2}, \cdots, y_{ij}, \cdots, y_{iN})$ is the practical output vector from the NN, corresponding to the sample set i;
N is the number of samples in each set;
P is the number of sample sets.

4.2.3 Validation process

In this step, the selected hidden-node scheme will be trained with new sample set (Input B in Figure 5) and a different sample set (Input C in Figure 5) will be selected to check the "best" hidden-layer scheme's validation. The purpose is to check the universality of the NN. The evaluation functions $G(W)$ and $G_{\max}(W)$ uses equations 4,5 and 6, except that the sample set changes and $P = 1$. If the values of $G(W)$ and $G_{\max}(W)$ accord with the criteria, the NN model with such hidden-node scheme will be applied.

Lastly, the validated NN can now be used to forecast future traffic data.

5. Empirical study

Given the complexity of the problem, three hidden-node schemes were designed, with 10, 20 and 30 nodes in the hidden layer.

5.1 Collecting sample sets

In the Chinese railway system, the main transportation task is planned and assigned on a monthly basis. Six east China sub-bureaux were selected and network data traffic from TMIS for four years (997 to 2000) were collected. The corresponding data listed in Table 3 on the transport task, was also collected. Each data set consisted of 48 points (monthly for four years), giving a total of 6 data sets for each of the six sub-bureaux.

5.2 Estimation Process

Firstly, three sample sets, which can be seen as Input A in Figure 5, were selected randomly and input to the NN. After the training and estimation processes, the evaluation functions were estimated and are shown in Table 4. The criteria used for the two functions are also shown in Table 4 for comparison. As $G(W)$ values for the three hidden-node schemes are all higher than the criteria values, the result cannot be accepted. Any two of the three sample sets used were combined into one large set. Another three "large" sample sets were formed to train each of three schemes. The estimation results are given in Table 5. The scheme with 30 nodes meets performs best at meeting the demand. In order to reach the stated level of precision enough number of samples are necessary to train NN. The hidden-node scheme with 30 nodes was used in the application described below.

5.3 Validation process

In this step, we trained the NN using a combination of randomly selected five sample sets, which can be seen as Input B in Figure 5. The NN hidden-node scheme was then tested with the last sample set, which can be seen as Input C in Figure 5. This last sample set was not used to calibrate the NN weights. This process will further demonstrate whether there exists a stable mapping relationship between input and output and whether NN can correctly identify the mapping relationship. In this procedure, the evaluation criteria are the same as those applied in estimation process.

The actual and the NN forecast values, are compared in Figure 6. The error distribution diagram is shown in Figure 7 and error statistical data is given in Table 6. It was found that the forecast result can adequately satisfy the criteria. It was demonstrated that the NN trained with data from one set of regions, can be used to forecast traffic data for other regions.

5.4 Forecasting process

The NN model was used to forecast the TMIS network data traffic of a railway sub-bureau[1] (Q) in the west of China. The input values for the NN were defined according to the planning construction scale and railway traffic in the next five years. This data was available only on an annual basis. We estimated the average monthly data for the next five years, shown here as input E, in Figure 5.

The NN model was trained with the entire six sample sets together. The combination of these six sample sets can be seen as Input D in Figure 5. The average monthly network data traffic of the future five years have been forecast as shown in Figure 8. After comparing the five

[1] The name of the sub-bureau has been withheld at this stage.

average monthly network data traffic estimates, the annual rate of increase in TMIS data traffic was estimates as shown in Figure 9.

5.5 Analysis of Result

As shown in Figure 8, the TMIS network data traffic of sub-bureau Q is increasing, for the next five years. In 2007, the data traffic will reach 1350 units[2] monthly, which is three times the 2002 value. As shown in Figure 9, the rate of increase reaches a maximum in 2005. According to the forward planning program for sub-bureau Q, in 2003 and 2004 parts of a new railway and TMIS engineering are being implemented or are under test running. By 2005, the transport task on the newly built lines will just fully begin operating regularly.. From this point, transportation production output may play a larger part on network data traffic than transportation equipment scale.

The network data traffic of a railway sub-bureau in east China, was estimated to have increased at a rate of 4%~5% annualy. This is due to the fact that the eastern railway network in China has not been extended in the recent past and capacity is close to saturation. However, with the historical railway TMIS data of easatern Chinas the sample sets used to train the NN, it was possible to produce reasonable forecasts for the west China network. This demonstrates the NN model is able to adequately reflect the mapping relationship.

6. Conclusion

The MoR in China has published its strategic goal in "the Tenth Five Years' Plan", which requires extending the main import/export corridors; increasing the density of the network in western China; enhancing construction techniques; and the reconstruction and improvement of the transport production capability. According to that plan, there will be 127 billion RMB Yuan invested on railway construction of western China. To the end of 2005, the length of western railway will reach 27.5 thousand kilometers, which accounts for 37 percent of the entire network in China. At that time the framework of western railway network will be basically in place.

There are two main tasks for the western China railways, namely: the extension of the network; and the modernization of the transportation equipment, including TMIS engineering. Only if these two tasks are successfully completed can the entire railway network perform its intended function effectively and efficiently. This paper offers an efficient method to forecast the network data traffic. The influences on network data traffic coming from transportation equipment scale and transportation production output .have been considered. In this way, the planner can clearly grasp the requirement arising from a practical railway transportation production.

There are some issues which require further study and research. Although the NN model has basically satisfied the demand, the maximum absolute error, $G_{\max}(W)$, has reached near 17%. This figure should be reduced in the future. It is possible that other influencing factors not taken into account here, also play a part. The NN model should be compared with other models or methods, even though the influence factors are very complex. Due to time constraints, the number of test samples was limited. It would be advisable to increase the number of data points available for the learning process.

[2] The unit of data traffic has not been licensed to publish in this paper by the project.

Reference

[1] Anantaram Balakrishnan, Thomas L. Magnanti, Richard T. Wong. (1995). "A decomposition algorithm for local access telecommunications network expansion planning, " Operations research 43, 58-76.
[2] Ben Yuhas, Nirwan Ansari. *Neural networks in telecommunications*. Boston: Kluwer Academic Publishers, 1994.
[3] Bruce Curry, Peter Morgan, Mick Silver. (2002). "Neural networks and non-linear statistical methods: an application to the modelling of price-quality relationships, ". Computers & Operations Research 29, 951-969.
[4] David Wright. (1998). "Analysis of the market for access to broadband telecommunications in the year 2000, " Computers & Operations Research 25, 127-138.
[5] Haykin, Simon S. *Neural networks: a comprehensive foundation*. Upper Saddle River, N.J. : Prentice Hall, 1999.
[6] Jaques Reifman, Earl E. Feldman. (2002). "Multilayer perceptron for nonlinear programming, " Computers & Operations Research 29, 1237-1250.
[7] Ljung, Lennart. (1999). *System identification: theory for the user*. Upper Saddle River, N.J.: Prentice Hall.
[8] M. Mozolin, J.C. Thill, E. Lynn Usery. (2000). "Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation, ". Transportation Research Part B 34, 53-73.
[9] Mark T. Leung, An-Sing Chen, Hazem Daouk. (2000). "Forecasting exchange rates using general regression neural networks, " Computers & Operations Research 27, 1093-1110.
[10] Nevil Brownlee. (1998). "Network Management and Realtime Traffic Flow Measurement, " Journal of Network and Systems Management 6, 223-228.
[11] Smith, Murray. (1993). *Neural networks for statistical modeling*. New York: Van Nostrand Reinhold.
[12] Thomas L. Magnanti, Prakash Mirchandani, Rita Vachani. (1995). "Modeling and solving the teo-facility capacitated network loading problem, " Operations Research 43, 142-157.
[13] Xing W. and Xie J.. (1999). *Modern optimization algorithms*. Beijing, P.R.China: Tsinghua Press.
[14] Zhang Yang Ming, (2001). "Grand Blueprint of Railway Development in the Tenth Five Year Plan: Visiting the director general of development & planning bureau of China MoR," Railway Knowledge of China 2, 4-10.